Different modes of stop codon restriction by the *Stylonychia* and *Paramecium* eRF1 translation termination factors

Sergey Lekomtsev*†, Petr Kolosov*, Laure Bidou†‡, Ludmila Frolova*, Jean-Pierre Rousset†‡, and Lev Kisselev*§

*Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow 119991, Russia; †Unité Mixte de Recherche 8621, Institut de Génétique et Microbiologie, Université Paris-Sud, F-91405 Orsay, France; and ‡Centre National de la Recherche Scientifique, F-91405 Orsay, France

Communicated by Dieter Söll, Yale University, New Haven, CT, April 28, 2007 (received for review December 8, 2006)

In universal-code eukaryotes, a single-translation termination factor, eukaryote class-1 polypeptide release factor (eRF1), decodes the three stop codons: UAA, UAG, and UGA. In some ciliates, like *Stylonychia* and *Paramecium*, eRF1s exhibit UGA-only decoding specificity, whereas UAG and UAA are reassigned as sense codons. Because variant-code ciliates may have evolved from universal-code ancestor(s), structural features should exist in ciliate eRF1s that restrict their stop codon recognition. In omnipotent eRF1s, stop codon recognition is associated with the N-terminal domain of the protein. Using both *in vitro* and *in vivo* assays, we show here that chimeric molecules composed of the N-terminal domain of *Stylonychia* eRF1 fused to the core domain (MC domain) of human eRF1 retained specificity toward UGA; this unambiguously associates eRF1 stop codon specificity to the nature of its N-terminal domain. Functional analysis of eRF1 chimeras constructed by swapping ciliate N-terminal domain sequences with the matching ones from the human protein highlighted the crucial role of the tripeptide QFM in restricting *Stylonychia* eRF1 specificity toward UGA. Using the site-directed mutagenesis, we show that *Paramecium* eRF1 specificity toward UGA resides within the NIKS (amino acids 61–64) and YxCxxxF (amino acids 124–131) motifs. Thus, we establish that eRF1 from two different ciliates relies on different molecular mechanisms to achieve specificity toward the UGA stop codon. This finding suggests that eRF1 restriction of specificity to only UGA might have been an early event occurring in independent instances in ciliate evolutionary history, possibly facilitating the reassignment of UAG and UAA to sense codons.

ciliated protozoa | dual gene reporter system | eukaryote class-1 polypeptide release factors | interdomain and intradomain protein chimeras | stop codon decoding

In the universal genetic code, three stop codons (UAA, UAG, and UGA) located at the termini of mRNA sequences are decoded at the termination step of translation by class-1 polypeptide release factors (RF) (reviewed in refs. 1–3). However, in organisms with variations in the genetic code, like ciliates, class-1 factors are able to decode only one or two stop codons with the remaining stop codon(s) reassigned to encode certain amino acids (for review, see refs. 4 and 5). The molecular mechanisms that restrict stop codon recognition are entirely unknown and represent major unresolved problems in molecular biology and genetics. In eukaryotes with the standard code, a single class-1 RF, designated eukaryote class-1 polypeptide release factor (eRF1), decodes all three stop codons. Stop codon decoding results in signal transduction from the small to the large ribosomal subunit leading to cleavage of peptidyl-tRNA at the peptidyl transferase center of the ribosome.

The eRF1 protein family is highly conserved and similar to archaeal class-1 RFs but differs profoundly from bacterial class-1 RFs (6–8). The only known structural element common to all class-1 RFs is a universal GGQ tripeptide essential for triggering peptidyl-tRNA hydrolysis at the peptidyl transferase center (8–15). Because of the absence of a common ancestor and the significant differences in domain organization and three-dimensional structure (8, 16, 17), it seems likely that the mechanisms used to decode stop codons in the ribosome by bacteria are different from those found in archaea and eukaryotes. In eukaryotes, the N-terminal domain of eRF1 is implicated in decoding, as shown by genetic (18), biochemical (19–23), and bioinformatic (24–26) data. More specifically, in human eRF1 (Hs-eRF1), the invariant dipeptide IK (positions 62–63) is likely to be involved in the recognition of the first nucleotide in the stop codon (21), whereas the YxCxxxF motif (positions 125–131) is implicated in purine discrimination at the second and third stop codon positions (20, 23).

It was shown that mammalian ribosomes are able to accommodate eRF1 from some variant-code organisms like the ciliate *Euplotes.* In this heterologous system, *Euplotes* eRF1 (Eu-eRF1) retains *in vitro* UAR-only stop codon specificity (27). Furthermore, a chimeric eRF1 composed of the N-terminal domain of *Tetrahymena* and the MC domain of yeast *Schizosaccharomyces pombe* exhibits the UGA-only specificity *in vitro* with the mammalian ribosomes (28). From these experiments, one can infer that the stop-codon decoding specificity, at least in some ciliates, is associated with the N-terminal domain of eRF1, as in the omnipotent eRF1s from universal-code organisms.

Because the evolutionary distance between mammals (the source of the ribosomes for our *in vitro* assay of RF activity) and ciliates (the source of eRF1s) is large, the ribosome-binding sites of ciliate eRF1s may not interact properly with the mammalian ribosomes. Although Eu-eRF1 is active with rabbit ribosomes (27), *Tetrahymena* eRF1 is not (29). To overcome this problem, we compared the decoding properties of various eRF1s by constructing interdomain chimeras in which the N and MC domains were derived from ciliate and human eRF1s, respectively. Mammalian ribosomes bind the MC domain of Hs-eRF1 in the absence of the N-terminal domain (30). A similar approach was described earlier for molecular chimeras of *Tetrahymena* and *S. pombe* eRF1 domains (28).

In this work, the MC domain of Hs-eRF1 was fused with the N-terminal domain of *Stylonychia* eRF1 (St-eRF1), which exhibits UGA-only specificity (5). The *Stylonychia*–human interdomain chimera appeared to be active in an *in vitro* RF assay. It was therefore possible to develop a strategy for the identification of stop-codon specificity discriminators via preparation and functional analysis of intra-N-terminal domain chimeras with swapped fragments from the N-terminal domains of the UGA-only St-eRF1 and of the omnipotent Hs-eRF1. Then, RF activity of intra-N-terminal

Author contributions: S.L. and P.K. contributed equally to this work; L.F. designed research; S.L. and P.K. performed research; L.B. contributed new reagents/analytic tools; L.B., J.-P.R. and L.K. analyzed data; and J.-P.R. and L.K. wrote the paper.

The authors declare no conflict of interest.

Abbreviations: RF, release factor; eRF1, eukaryote class-1 polypeptide RF; Eu-eRF1, *Euplotes* eRF1; Hs-eRF1, human eRF1; Pa-eRF1, *Paramecium* eRF1; St-eRF1, *Stylonychia* eRF1.

§To whom correspondence should be addressed. E-mail: kissel@eimb.relarn.ru.

This article contains supporting information online at www.pnas.org/cgi/content/full/0703887104/DC1.

domain chimeras was assayed *in vitro* to determine the type of stop codon specificity for the given constructs.

An *in vitro* RF activity assay (31) is suitable for comparison of different stop codon responses of various eRF1s (6, 9, 21, 23). However, this assay lacks natural mRNA, class-2 release factor, eRF3, GTP, and natural peptidyl-tRNA. We complemented the *in vitro* assay with an *in vivo* measurement of stop codon readthrough by using a dual gene reporter system (32, 33). Although both assays are essentially distinct, they provided similar results. Together these assays validate the alteration of stop codon specificity in the chimeric eRF1s.

The goal of this work was to identify amino acid residues that confer UGA-only specificity to St-eRF1 by using a molecular chimera approach followed by *in vitro* functional analysis and *in vivo* readthrough level measurement. In addition, some amino acid residues of the N-terminal domain specific for the UGA-only *Paramecium* eRF1 (Pa-eRF1) were introduced into respective positions of Hs-eRF1 to confer the UGA-only specificity. In ciliate eRF1s with the same UGA-only specificity, we identified the amino acid residues that limit stop codon recognition of these eRF1s to UGA.

Results

Interdomain eRF1 Chimeras Possess Stop Codon Specificity Defined by Their N-Terminal Domains. In organisms with the universal genetic code, stop-codon specificity is associated solely with the N-terminal domains of eRF1s. It is less clear whether the N-terminal domain also is associated with stop codon specificity in organisms with variant codes. For example, for Eu-eRF1, cross-linking (22) and genetic (34) data are consistent with a dominant role for the N-terminal domain in stop codon discrimination. In contrast, for *Tetrahymena* eRF1, the data are controversial. Chimeric constructs composed of the N-terminal domain of *Tetrahymena* eRF1 fused with the MC domain of yeast eRF1 is omnipotent *in vivo* (34), but this chimera only recognizes UGA *in vitro* (28). Several hypothetical explanations were suggested to account for this apparent discrepancy (34).

We used two assays to evaluate the functional activity of eRF1s. The first assay (31) explores a simplified system in which the role of peptidyl-tRNA is mimicked by fMet-tRNA; two oligoribonucleotides are used instead of natural mRNA. The two oligoribonucleotides include an AUG that binds to fMet-tRNA at the ribosomal P site and one of the three stop-codon-containing tetranucleotides, UAAA, UAGA, or UGAA, that bind to the A site. In the second assay, *in vivo* stop codon readthrough in intact cells is quantified as an inverse measurement of RF activity (32, 33).

To examine the role of the N-terminal domain of eRF1s from variant-code organisms, we prepared interdomain chimera between the N-terminal domain of the UGA-only St-eRF1 and the MC domain of omnipotent Hs-eRF1 (Fig. 1, St-1). RF activity and readthrough measurements showed that, in St-eRF1, stop codon discrimination is tightly associated with the N-terminal domain, as in eRF1s from universal-code organisms. The MC domain of omnipotent Hs-eRF1 has no effect on the decoding specificity of the chimeric St-eRF1 (Fig. 1).

RF Activity of *Stylonychia* and Human eRF1 Intra-N-Terminal Domain Chimeras. In interdomain constructs, stop codon specificity is governed by their N-terminal domains, which opened a way for creating intra-N-domain chimeras composed of sequences derived both from UGA-only ciliate St-eRF1 and from omnipotent Hs-eRF1.

In construct St-2, the N-terminal part (positions 1–82) was derived from Hs-eRF1, whereas in positions 83–145 a sequence from the St-eRF1 N-terminal domain was placed (Fig. 1). This chimera possessed mostly the UGA-only specificity, demonstrating that the determinant(s) for the UGA-only recognition resides in the C-terminal moiety of the N-terminal domain. Construct St-3 is

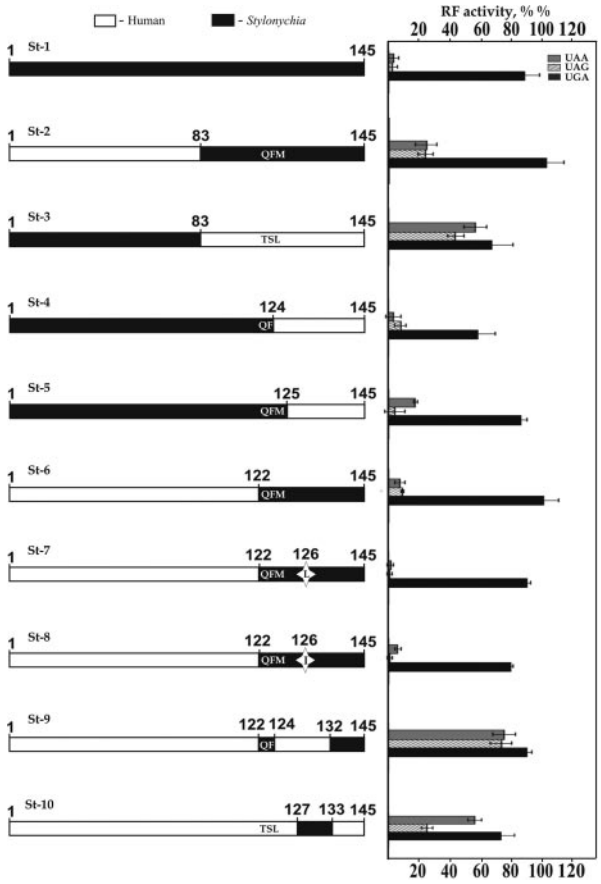

Fig. 1. *In vitro* RF activity of chimeric eRF1 constructs containing the whole N-terminal domain of St-eRF1 (St-1) and swaps between human and *Stylonychia* sequences within the N-terminal domains (St-2 to St-10). All eRF1 constructs contained the MC domain of Hs-eRF1 (positions 145–437). The junctions between the N and MC domains corresponding to amino acids 144 and 145 of Hs-eRF1 are located in a hinge region between the N and M domains. The numeration of amino acid positions in all constructs is as in the WT-Hs-eRF1. TSL sequence (positions 122–124) from Hs-eRF1 and QFM sequence at the same positions from St-eRF1 are indicated. The numbers on the border of swapped sequences correspond to the first amino acid residue of the right-hand sequence.

reciprocal to St-2 because it begins with an St-eRF1 sequence (positions 1–82) followed by the human sequence (positions 83–145). St-3 responded to all three stop codons, with only a 2-fold reduction in RF activity. Thus, in St-eRF1, stop-codon specificity determinants are located between positions 83 and 145. Construct St-4, Hs-eRF1 (positions 124–145), and St-eRF1 (positions 1–123) chimera remained predominantly specific for UGA decoding, although the response was reduced compared with St-1. In construct St-5, the border between the human and *Stylonychia* sequences was shifted downstream by one amino acid residue compared with St-4. As a result, the UGA response increased from 60% to 90%. Construct St-6, which contained human (positions 1–121) and *Stylonychia* (positions 122–145) sequences, was clearly specific for UGA.

Functional analysis of the St-4, St-5, and St-6 chimeras pointed to an essential role of positions 122–124 in stop codon discrimination. In fact, the UGA-only specificity was predominantly associated with amino acids Q122 and F123 but was also partly affected by position 124 because an L124M substitution increased the UGA response (Fig. 1).

As is evident from an alignment of the amino acid sequences, some positions in the 122–131 fragment are variable, whereas others are highly conserved (Fig. 2). Based on genetic (18) and

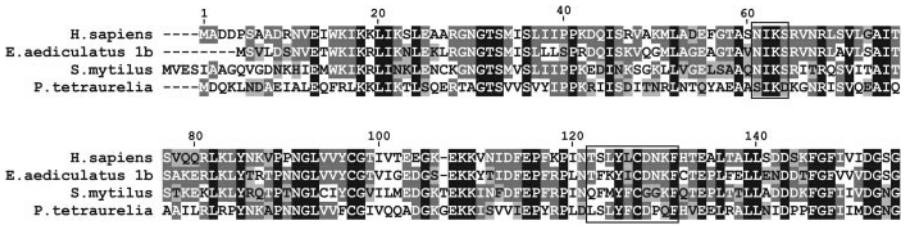

Fig. 2. Amino acid sequences from the N-terminal domains of omnipotent (human), unipotent (*Stylonychia* and *Paramecium*), and bipotent (*Euplotes*) eRF1s. Residue numbering is as in the WT-Hs-eRF1. Identical, conserved, and semiconserved amino acid residues are black, dark gray, and light gray, respectively.

bioinformatic analyses (26), position 126, which is differentially conserved in omnipotent (L126), UAR-specific (I126), and UGA-specific (F126) eRF1s, has been suggested as potentially important for stop codon discrimination. However, in Hs-eRF1, substitutions of L126 had no effect on RF activity *in vitro* (23). In constructs St-7 (F126L as in Hs-eRF1) and St-8 (F126I as in Eu-eRF1), these substitutions were introduced into the *Stylonychia* context (positions 122–145). In both constructs, the UGA-only response was completely retained (Fig. 1). Thus, presumably, the amino acid identity at position 126 is insignificant for St-eRF1 decoding specificity in the *Stylonychia* context and for Hs-eRF1 specificity in the human context.

***In Vivo* Analysis of the Intra-N-Terminal Domain Chimeras Between *Stylonychia* and Human eRF1s.** A dual gene reporter system (32, 33) was applied to quantify the effect of chimeric eRF1 on stop codon readthrough in cultured human cells (Fig. 3). The C293 cells were cotransfected with one of the test vectors (St-1, St-6, and St-11–St-15) and one of the reporter vectors. The latter contained one of the three stop codons between the *lac*Z and *luc* genes. The readthrough level at each stop codon was measured. In control experiments, to measure the effect of expression of exogenous WT Hs-eRF1 on readthrough level, cells were transfected with reporter vector and pcDNA3 or pcDNA3 bearing the WT-Hs-eRF1 gene (pcDNA3-Hs-eRF1). The readthrough level, which differed insignificantly, was low for pcDNA3 (0.3%, 1.7%, and 0.7% at UAA, UAG, and UGA, respectively) and pcDNA3-Hs-eRF1 (0.3%, 2.0%, and 0.8% at UAA, UAG, and UGA, respectively). The readthrough levels for cells transfected with the WT-Hs-eRF1 gene and for cells transfected with chimeric genes were compared. To determine a value for maximum (100%) readthrough, assays were done with the reporter vector that contained a sense codon in place of the stop codon. For all experiments, we used a calcium phosphate DNA coprecipitation protocol to favor cotransfection events in which a single cell received both the reporter and test vector DNAs.

We found (Fig. 4*B*) that the chimeric construct with the entire *Stylonychia* N-terminal domain (St-1) exhibited at the UGA stop codon nearly the same (1.4-fold difference) level of readthrough (1.1%) as did the control construct with the entire human sequence (0.8%). In contrast, the levels of readthrough observed at UAA (6.0%) and UAG (11.3%) stop codons were considerably higher than that of the readthrough in the presence of the WT-Hs-eRF1. These values were 18.9-fold (UAA) and 5.5-fold (UAG) higher than the level of readthrough allowed by the WT-Hs-eRF1. These data were in full agreement with *in vitro* experiments (Fig. 4*A*) and with the fact that St-eRF1 only recognizes UGA as a stop codon (5).

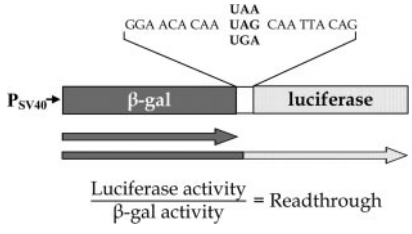

Fig. 3. The scheme for the dual gene reporter system used to monitor translation readthrough *in vivo*.

Construct St-6, containing human (positions 1–121) and *Stylonychia* (positions 122–145) sequences, was clearly specific for UGA *in vitro* (Figs. 1 and 4*A*). When St-6 was expressed in human cells (Fig. 4*B*), we observed higher levels of readthrough at UAA (6.1%, 19.1-fold increase) and UAG (11.0%, 5.4-fold increase) in contrast to UGA (1.1%, 1.4-fold increase). These results suggest that the positions 122–145 of St-eRF1 are capable of discriminating UAA and UAG stop codons both *in vitro* and *in vivo*.

In vitro, the UGA-only specificity was observed for St-11, whereas omnipotent specificity was shown for St-12 (Fig. 4*A*). We found that St-11 exhibited a high level of readthrough at UAA (5.3%, 16.7-fold increase) and UAG (13.9%, 6.8-fold increase) and a low level at UGA (1.3%, 1.7-fold increase). However, readthrough was significantly lower at UAA (1.2%, 3.8-fold increase) and UAG (6.4%, 3.1-fold increase) in cells expressing St-12. The readthrough level for St-12 at UGA (2.4%) increased 3.1-fold (Fig. 4*B*). These results indicate that residues in positions 132–145 are not involved in stop codon discrimination both *in vitro* and *in vivo*.

The QFMYF, QFM, and QF sequences (positions 122–126) from St-eRF1 were inserted into the human context (St-13, St-14, and St-15, respectively). The St-13 chimera appeared to be of the UGA-only type *in vitro*. At the same time, St-13 exhibited a high level of *in vivo* readthrough at UAA (3.6%, 11.2-fold increase) and UAG (11.9%, 5.8-fold increase) but a low level at UGA (1.1%, 1.4-fold increase) as compared with the basal level of readthrough allowed by the WT-Hs-eRF1. In contrast, St-14 was somewhat omnipotent *in vitro*. Readthrough enhancement was lower at UAA (1.2%, 3.9-fold increase) and at UAG (7.8%, 3.8-fold increase) than for St-13. The readthrough value for St-14 at UGA (1.7%, 2.2-fold increase) was not significantly different from the same value for St-13. The St-15 was omnipotent *in vitro* and showed 2.0% readthrough at UAA, 9.0% at UAG, and 2.5% at UGA, with 6.2-, 4.4-, and 3.1-fold increases of readthrough level, respectively. Thus, neither QF nor QFM sequence is sufficient to bring the stringent UGA-only specificity to the human context, and the presence of F126 is required for the UGA-only specificity.

The combination of the QF dipeptide with the human amino acid sequence (positions 124–131) led to the appearance of responses to UAA and UAG *in vitro*, although they were weaker than the UGA response (construct St-9). Omnipotent specificity was shown for St-10, where the QFM tripeptide was absent (Fig. 1).

Taken together, these results (Fig. 4) imply that positions 122–126 of the *Stylonychia* N-terminal domain are crucially important for the UGA-only specificity both *in vitro* and *in vivo*.

Stop Codon Restriction in *Paramecium* eRF1. As is evident from the alignment (Fig. 2) of the UGA-only St-eRF1 and Pa-eRF1, the variable amino acid residues in positions 122–131 are different. In Pa-eRF1, the LSL tripeptide (positions 122–124), which resembles the human TSL rather than the *Stylonychia* QFM tripeptide, is present. At the same time, the GGK tripeptide of St-eRF1 (positions 128–130) is replaced with the DPQ tripeptide in Pa-eRF1, whereas in Hs-eRF1 the DNK tripeptide occurs at the same place (Fig. 2).

We introduced the *Paramecium* sequence (positions 122–131)

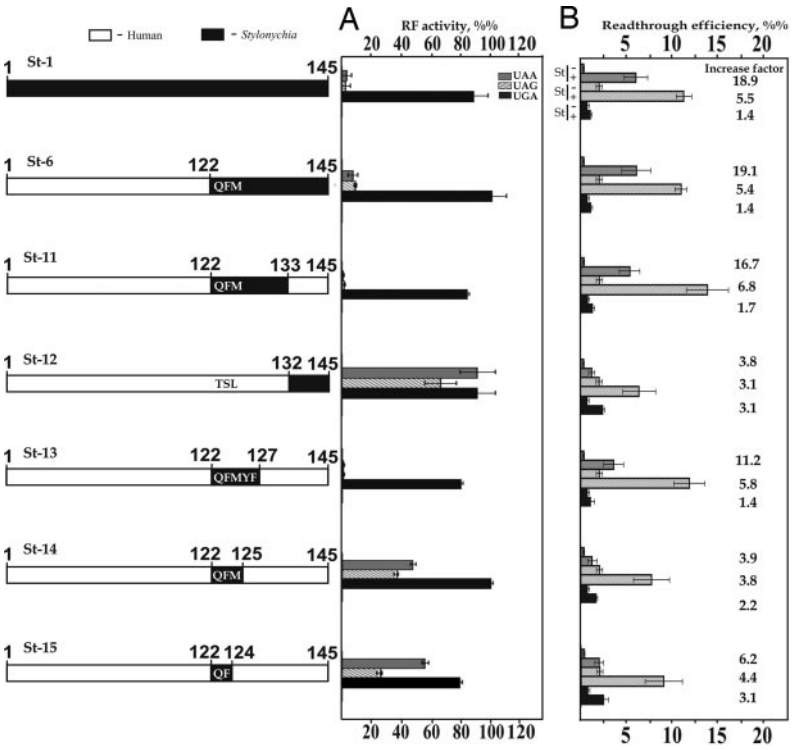

Fig. 4. Measurements of RF activity *in vitro* (*A*) and stop codon readthrough in dual reporter system (*B*). (*A*) For details of *in vitro* RF assay, see *Methods*. (*B*) Readthrough levels for each of the stop codons were measured in the presence of the WT-Hs-eRF1 (−) or the chimeric (+) eRF1s. All eRF1 constructs contained the MC domain of Hs-eRF1 (positions 145–437).

into Hs-eRF1 (Fig. 5, Pa-1) and observed omnipotent stop codon recognition. The omnipotent stop codon response also was demonstrated for the N61S+S64D double Hs-eRF1 mutant, but a diminution of RF activity toward all three stop codons was observed (Fig. 5, Pa-3). This mutant has substitutions corresponding to the Pa-eRF1 sequence in this region (20). When introduced separately, these substitutions are thus insufficient to bring the UGA-only specificity to Hs-eRF1. This result is in sharp contrast to the case of *Stylonychia*, in which the pentapeptide 122–126 provided the UGA-only specificity to the human context. The N61S and S64D mutations in the NIKS motif, which correspond to the Pa-eRF1 amino acid sequence in this region, endowed Pa-1 with UGA-only specificity (Fig. 5, Pa-2). The same UGA-only specificity, but with a decreased UGA response compared with Pa-2, was shown for the quadruple mutant N61S+S64D+N129P+K130Q of Hs-eRF1 (20). All these results suggest that the Pa-eRF1 discriminator sites are predominantly located inside the NIKS (positions 61 and 64) and YxCxxxF (positions 129 and 130) motifs.

Discussion

Functional Assays for Measuring Decoding Specificity of eRF1s. To evaluate the stop-codon-dependent translation termination activity of RF1s, the first assay (31) explores a system in which peptidyl-tRNA is substituted by fMet-tRNA, AUG binds to fMet-tRNA at the P site, and tetraplets UAAA, UAGA, or UGAA bind to eRF1 at the A site. In the second assay, readthrough of stop codons in intact cells is measured (32, 33). Eukaryotic protein-synthesizing machinery fully reconstituted *in vitro* (35) can also be used to measure eRF1 stop codon specificity and RF activity.

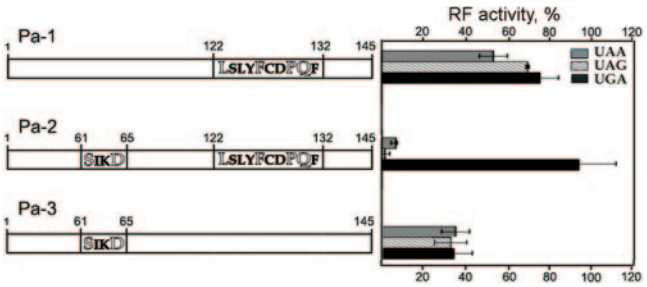

Fig. 5. RF activity of the human eRF1 mutants with substitutions in the N-terminal domain corresponding to the amino acid sequence of Pa-eRF1. Constructs Pa-1 and Pa-2 contained the MC domain of Hs-eRF1 (positions 145–437). Pa-3 is a full-length Hs-eRF1 with substitutions only in the NIKS sequence (data from 20). For details, see *Methods*.

For the aims of these studies, we selected the first two systems because of their simplicity. Both *in vitro* and *in vivo* systems are particularly suitable for the comparison of different mutant and chimeric eRF1 responses toward different stop codons. The absence of natural mRNA, eRF3, and GTP/GDP in the first assay may influence the response ratio of eRF1 to different stop codons. To exclude this possibility, we have applied a dual gene reporter system to the evaluation of eRF1 decoding specificity in mammalian cells (32, 33) that measures readthrough efficiency in the presence of transfected exogenous chimeric or mutant eRF1 and endogenous Hs-eRF1 for the different stop codons. It is known that eRF1 binds strongly to the ribosome because of its interaction with a stop codon at the A site (36). Therefore, if the efficiency of chimeric or mutant eRF1 binding to a given stop codon is strongly reduced, the ability of that eRF1 to compete with tRNA(s) and/or endogenous eRF1 for that stop codon is significantly diminished. Such a weakly binding eRF1 facilitates the binding of near-cognate and/or suppressor tRNA(s) to the A site, resulting in an increase of readthrough level at that stop codon. In controls, where the reporter vector was cotransfected with a vector that allowed expression of the entire human eRF1, the readthrough level of all three stop codons was very low as anticipated (Fig. 4*B*).

The readthrough data fully confirmed the *in vitro* results. Constructs lacking the ability to respond to UAA and UAG *in vitro* (St-1, St-6, St-11, and St-13) are unable to compete with tRNA and a high readthrough level is the result (Fig. 4*B*). In contrast, for constructs St-12, St-14, and St-15, which retain their omnipotent response *in vitro* (Fig. 4*A*), the readthrough level is low because binding to the stop codons in the ribosome is retained (Fig. 4*B*).

Because the *in vivo* experiments have been performed in the presence of endogenous Hs-eRF1, eRF3, GTP/GDP, and tRNAs,

these results demonstrate that the chimeric and mutant eRF1s are "dominant" over the endogenous Hs-eRF1. This is likely due to the overexpression of the chimeric and mutant eRF1s as directed by the strong cytomegalovirus promoter that would titrate eRF3 and prevent it from interacting with endogeneous WT-Hs-eRF1. It is worth mentioning that expression of exogenous Hs-eRF1 (Fig. 4*B*) negligibly affects the basal readthrough level. This may imply that basal readthrough efficiency primarily depends on the mRNA context and/or concentrations of suppressor and near-cognate tRNAs rather than on excessive amounts of eRF1.

Thus, the absence of eRF3/GTP/GTP in an *in vitro* assay and the presence of endogenous eRF1 in an *in vivo* assay do not interfere with specificity of stop codon recognition by chimeric eRF1s.

Discriminator Sites in the UGA-Only *Stylonychia* and *Paramecium* eRF1s Are Profoundly Different and Located in the N-Terminal Domains. We have shown that stop codon discrimination is governed by the N-terminal domain of the unipotent St-eRF1 even when fused with the MC domain of the omnipotent Hs-eRF1. Construct St-1 (Fig. 1) is completely active with UGA and possesses very low (if any) activity in response to UAA and UAG codons as based on the variant genetic code of this ciliate. Furthermore, genetic *in vivo* and *in vitro* data support the view that stop codon discrimination is entirely associated with the N-terminal domain of Eu-eRF1 (22, 34). Both previous (20) and our present data (Fig. 5) make it likely that in Pa-eRF1 stop codon discrimination is also governed by the N-terminal domain. Consequently, in both UGA-only eRF1s and universal-code human and yeast eRF1s, stop-codon specificity ultimately depends on the N-terminal domain structure. However, it remains unclear whether the "N-terminal domain rule" is applicable to all variant-code ciliates because at present the data regarding stop-codon discrimination for *Tetrahymena* eRF1 are apparently controversial (28, 34) and require further studies.

Stylonychia and *Paramecium* species have the same UGA-only type of stop codon recognition (for review, see ref. 5). By comparing the decoding specificities of the N-terminal domains of eRF1s, it is possible to find out whether similar or distinct strategies are used by these ciliates to restrict recognition of two stop codons by eRF1s.

The UAA and UAG responses are eliminated in Hs-eRF1 by introducing the St-eRF1 pentapeptide QFMYF (positions 122–126) (Fig. 4). This *Stylonychia* sequence in the context of human N-terminal domain converts the omnipotent eRF1 into a unipotent one (St-13). If F126 is replaced by L126 from the human sequence, weak recognition for each stop codon reappears (St-14). On the other hand, human L126 in the *Stylonychia* context is unable to abolish the UGA-only response (St-7) (Fig. 1). The dominant role of the QF dipeptide in QFMYF is evident from the functional analysis of constructs St-4, St-5, and St-6, where QF keeps more or less the UGA-only response depending on the context. Consequently, we have identified the discriminator site, the QFMYF pentapeptide in which the major role is associated with QF, whereas M124 and F126 strongly affect the QF discrimination ability depending on the sequence context.

Given that the YxCxxxF motif (positions 125–131) is essential for purine discrimination in stop codons (20, 23), we assume that the decapeptide (positions 122–131) is the major stop codon recognition and discriminator site in Hs-eRF1 and St-eRF1.

In some constructs (for instance, St-2, St-4, St-5, and St-6), a low RF activity with UAA and UAG codons is observed, although these constructs are predominantly of the UGA-only type. We suggest that the presence of human sequences in St-eRF1 might relax the UGA-only specificity in the chimeric constructs.

Pa-eRF1 is also of the UGA-only type, but the 122–131 peptide from Pa-eRF1 inserted into the human context is unable alone to induce a UGA-only response, although the UAA and UAG responses are diminished (Fig. 5). When in Pa-1 the human NIKS sequence is replaced by Pa-eRF1 SIKD sequence (Pa-2), a complete conversion to the UGA-only type is observed. Earlier, the same result was achieved with Hs-eRF1 when positions 61, 64, 129, and 130 were replaced with residues from the Pa-eRF1 sequence (20).

In Pa-eRF1, the discriminators are tightly associated with the NIKS and YxCxxxF motifs essential for stop codon recognition, whereas in St-eRF1 only the extended part of the YxCxxxF motif (positions 122–126) is a discriminator. Molecular modeling showed that eRF1 can undergo a significant conformational change in the ribosome (37, 38), as suggested earlier (36). Because of this change, the NIKS and YxCxxxF motifs become proximal to the stop codon at the A site (38), which implies that, for these eRF1s, a probable mechanism for stop codon restriction is associated with an interaction in space between discriminator and recognition sites, leading to local conformational changes of the latter. This conformational change in the recognition site, in turn, results in a distortion of the recognition site binding to UAA and UAG, leaving the UGA binding intact. To prove this hypothesis, a cocrystal with the ribosome is needed to establish the exact structure of the N-terminal domains of ciliate eRF1s at the ribosomal decoding site.

General Remarks. Phylogenetic trees based on eRF1 sequences have been constructed and modified (4, 5, 25, 39). Despite some differences in these trees, it seems obvious that all known ciliate eRF1s diverged from the ancestral omnipotent eRF1 very early. Ciliate eRF1s, even with the same type of stop codon recognition like *Paramecium* and *Stylonychia*, are very distant on the phylogenetic tree and not clustered together (25). These data on ciliate eRF1 evolution are consistent with the results described above (Figs. 1, 4, and 5), and the structural constraints operating in their eRF1s are profoundly different, as shown in this work.

Molecular modeling of RNA-binding propensities for amino acid residues of ciliate and conventional eRF1s has been undertaken (25). RNA-binding propensities of eRF1s from universal-code organisms seem to be similar. In contrast, this pattern was different for various ciliate eRF1s. For example, one of the major differences in RNA-binding propensities was found for St-eRF1 in positions 123–126 that fit with the location of the St-eRF1 discriminator site (Figs. 1 and 4). Construct St-1 entirely lost its ability to recognize UAA and UAG (Figs. 1 and 4). UAR-only Eu-eRF1 does not cross-react with UGA; instead, UAR-only Eu-eRF1 retains its contact with the other two stop codons (22). These eRF1s are most likely unable to pass the codon-dependent step of eRF1 binding to the ribosome (36) and, consequently, do not compete with tRNAs for the "free" stop codons. For this reason, near-cognate mispairing between certain tRNAs and the free stop codon(s) may become a starting point for stop codon reassignment. Our data strongly argue in favor of the idea that *Stylonychia* and *Paramecium* followed distinct paths to diverge from a universal-code ancestor (5), as was suggested earlier for *Euplotes* and *Tetrahymena* (34).

Methods

Chimeric *Stylonychia mytilus*/Human eRF1 Constructions. The full-length double-stranded cDNA encoding Hs-eRF1 with a unique *Bst98I* site affecting neither amino acid sequence nor the reading frame was cloned into pET23b(+) vector (Novagen, San Diego, CA) as described (10). The resulting plasmid was designated pERF4b. For swapping experiments of the N-terminal domain sequences between human and *S. mytilus* eRF1s, additional HindIII and SalI restriction sites were introduced into pERF4b, corresponding to amino acid positions 83–84 and 144–145, respectively. The resulting plasmid was designated pERF4b-Sal.

The pERF4b-LD plasmid encoding eRF1 with two amino acid substitutions, S144L and K145D, arose after ligation of the amplified NdeI/XhoI DNA fragment encoding the N-terminal domain of Hs-eRF1 with pERF4b-Sal hydrolyzed with NdeI and SalI. All chimeric constructs contained these two amino acid substitutions, which did not affect the RF activity of eRF1.

The *eRF1* gene sequence encoding the N-terminal domain of *S.*

mytilus eRF1 was amplified by PCR from *S. mytilus* cell culture. The resulting PCR product was cloned into pET23b(+) vector at the NdeI/XhoI restriction sites. One intron is present in the ORF encoding the N-terminal domain of St-eRF1 (5). To create an intronless transcript, exon I amplified by PCR was cloned into pET23b by using NdeI/HindIII sites. Then, exon II amplified by PCR was cloned into plasmid with exon I at HindIII/XhoI sites. Five in-frame stop codons (TAA/TAG) present in the N-terminal domain of St-eRF1 were changed to CAA/CAG codons by using the PCR-based "megaprimer" method (40). This *S. mytilus* N-terminal domain fragment then was fused to a fragment encoding MC domain from human eRF1 in pERF4b-Sal, yielding the *S. mytilus*/human eRF1 expression plasmid.

To make plasmids with chimeric eRF1 constructs for *in vivo* transfection, human eRF1 gene was cloned into pcDNA3 (Invitrogen, Carlsbad, CA) at HindIII/XbaI sites. Then, the N-terminal part of Hs-eRF1 of the resulting pcDNA3–Hs-eRF1 plasmid was replaced by PCR products of N-terminal parts of the chimeric St-eRF1 constructs (St-1, St-6, and St-11–St-15) at HindIII/BamHI sites.

For information on swapping the N-terminal domain sequences between human and *S. mytilus* eRF1s, see supporting information (SI) *Materials and Methods*.

Insertion of *Paramecium tetraurelia* eRF1 Sequences into the N-Terminal Domain of Human eRF1. To construct Pa-1, mutations T122L+L126F+N129P+K130Q (residue numbering as in Hs-eRF1) were introduced into pERF4b by site-directed mutagenesis. Mutation N61S+S64D was introduced into the Pa-1 sequence, which yielded a plasmid-encoding Pa-2 construct.

Expression and Purification of eRF1s and Preparation of the Ribosomes. The WT-Hs-eRF1 and chimeric constructs containing His_6-tags at the C termini were produced in *Escherichia coli* strain BL21(DE3) and purified by using Ni-NTA Superflow (Qiagen, Valencia, CA) as described in ref. 19. Rabbit reticulocyte 80S ribosomes were isolated and purified as described in ref. 20.

***In Vitro* RF Activity Assay.** The eRF1 activity was measured as described (6, 31) at saturating levels of UAAA, UAGA, or UGAA stop-codon-containing tetraplets. The background measured without tetraplet was subtracted from all values. All measurements were repeated three times. The RF activity for the WT-Hs-eRF1 was considered as 100% with respect to the given stop codon. The RF activities of the chimeric or mutant eRF1s were calculated as percentages of the 100% RF activity value of the WT-Hs-eRF1 for the respective stop codon.

Cell Culture and Transfection. Human fibroblast C293 was incubated at 37°C in DMEM supplemented with 10% FCS in humidity-saturated 5.5% CO_2 in air. One million cells were cotransfected with 10 μg of each test vector and 10 μg of the reporter vector, using DNA/phosphate coprecipitation. Cells were harvested on day 3, and crude extracts were prepared as described in ref. 41. For each construct, four independent transfection experiments were performed.

Quantification of Readthrough Efficiency. One of the three stop codons in the tobacco mosaic virus RNA context was inserted in-frame into reporter vector (pAC) between the *lac*Z and *luc* coding sequences (Fig. 3) (32, 33). All translating ribosomes would thus result in β-galactosidase synthesis, but only those reading through the stop codon would yield an active luciferase. This dual gene reporter system provided an internal control for calibration of individual experiments for the overall expression level of each construct to take into account such factors as vector stability, transfection efficiency, transcriptional, and translational rates. Readthrough efficiency was estimated by the ratio of luciferase activity to β-galactosidase activity. To establish the relative activities of these enzymes when expressed in equimolar amounts, the ratio of luciferase activity to β-galactosidase from an in-frame control plasmid was taken as a reference. Readthrough frequency expressed as a percentage was calculated by dividing the luciferase/β-galactosidase ratio obtained from each construct by the same ratio obtained with the in-frame control construct (33). β-galactosidase and luciferase activities were assayed in the same crude extract as described in refs. 42 and 43.

We thank Yves Mechulam for *E. coli* methionyl-tRNA synthetase and formylase; Maria Rautian for *S. mytilus*; David Bedwell for making his results (34) available before publication; Richard D'Ari, Alex Ambrogelly, and Patrick O'Donoghue for help with the manuscript; and two referees for valuable comments and criticism. This work was supported by Presidential Program on Supporting the Leading Russian Schools in Science Grant 1227.2006.4 (to L.K.), the Molecular and Cellular Biology Program of the Presidium of the Russian Academy of Sciences (L.K.), Russian Foundation for Basic Research Grants 05-04-49385 (to L.F.) and 06-04-48037 (to L.K.), Association Française Contre les Myopathies Grants 9584 and 10683 (to J.P.R.), Association pour la Recherche sur le Cancer Grant 3849 (to J.P.R.), and INTAS Young Scientist Fellowship Grant 05-109-4722 (to S.L.).

1. Kisselev L, Ehrenberg M, Frolova L (2003) *EMBO J* 22:175–182.
2. Nakamura Y, Ito K (2003) *Trends Biochem Sci* 28:99–105.
3. Poole ES, Askarian-Amiri ME, Major LL, McCaughan KK, Scarlett DJ, Wilson DN, Tate WP (2003) *Prog Nucleic Acid Res Mol Biol* 74:83–121.
4. Inagaki Y, Doolittle WF (2001) *Nucleic Acids Res* 29:921–927.
5. Lozupone CA, Knight RD, Landweber LF (2001) *Curr Biol* 11:65–74.
6. Frolova L, Le Goff X, Rasmussen HH, Cheperegin S, Drugeon G, Kress M, Arman I, Haenni AL, Celis JE, Philippe M, *et al.* (1994) *Nature* 372:701–703.
7. Kisselev LL, Buckingham RH (2000) *Trends Biochem Sci* 25:561–566.
8. Song H, Mugnier P, Das AK, Webb HM, Evans DR, Tuite MF, Hemmings BA, Barford D (2000) *Cell* 100:311–321.
9. Frolova LY, Tsivkovskii RY, Sivolobova GF, Oparina NY, Serpinsky OI, Blinov VM, Tatkov SI, Kisselev LL (1999) *RNA* 5:1014–1020.
10. Seit-Nebi A, Frolova L, Justesen J, Kisselev L (2001) *Nucleic Acids Res* 29:3982–3987.
11. Zavialov AV, Mora L, Buckingham RH, Ehrenberg M (2002) *Mol Cell* 10:789–798.
12. Klaholz BP, Pape T, Zavialov AV, Myasnikov AG, Orlova EV, Vestergaard B, Ehrenberg M, van Heel M (2003) *Nature* 421:90–94.
13. Mora L, Heurgue-Hamard V, Champ S, Ehrenberg M, Kisselev LL, Buckingham RH (2003) *Mol Microbiol* 47:267–275.
14. Rawat UB, Zavialov AV, Sengupta J, Valle M, Grassucci RA, Linde J, Vestergaard B, Ehrenberg M, Frank J (2003) *Nature* 421:87–90.
15. Petry S, Brodersen DE, Murphy FV, IV, Dunham CM, Selmer M, Tarry MJ, Kelley AC, Ramakrishnan V (2005) *Cell* 123:1255–1266.
16. Vestergaard B, Van LB, Andersen GR, Nyborg J, Buckingham RH, Kjeldgaard M (2001) *Mol Cell* 8:1375–1382.
17. Shin DH, Brandsen J, Jancarik J, Yokota H, Kim R, Kim SH (2004) *J Mol Biol* 341:227–239.
18. Bertram G, Bell HA, Ritchie DW, Fullerton G, Stansfield I (2000) *RNA* 6:1236–1247.
19. Frolova L, Seit-Nebi A, Kisselev L (2002) *RNA* 8:129–136.
20. Seit-Nebi A, Frolova L, Kisselev L (2002) *EMBO Rep* 3:881–886.
21. Chavatte L, Seit-Nebi A, Dubovaya V, Favre A (2002) *EMBO J* 21:5302–5311.
22. Chavatte L, Kervestin S, Favre A, Jean-Jean O (2003) *EMBO J* 22:1644–1653.
23. Kolosov P, Frolova L, Seit-Nebi A, Dubovaya V, Kononenko A, Oparina N, Justesen J, Efimov A, Kisselev L (2005) *Nucleic Acids Res* 33:6418–6425.
24. Inagaki Y, Blouin C, Doolittle WF, Roger AJ (2002) *Nucleic Acids Res* 30:532–544.
25. Kim OT, Yura K, Go N, Harumoto T (2005) *Gene* 346:277–286.
26. Liang H, Wong JY, Bao Q, Cavalcanti AR, Landweber LF (2005) *J Mol Evol* 60:337–344.
27. Kervestin S, Frolova L, Kisselev L, Jean-Jean O (2001) *EMBO Rep* 2:680–684.
28. Ito K, Frolova L, Seit-Nebi A, Karamyshev A, Kisselev L, Nakamura Y (2002) *Proc Natl Acad Sci USA* 99:8494–8499.
29. Karamyshev AL, Karamysheva ZN, Ito K, Matsufuji S, Nakamura Y (1999) *Biochemistry (Moscow)* 64:1391–1400.
30. Dubovaya VI, Kolosov PM, Alkalaeva EZ, Frolova L, Kiselev LL (2006) *Mol Biol (Moscow)* 40:310–316.
31. Caskey CT, Beaudet AL, Tate WP (1974) *Methods Enzymol* 30:293–303.
32. Stahl G, Bidou L, Rousset JP, Cassan M (1995) *Nucleic Acids Res* 23:1557–1560.
33. Bidou L, Hatin I, Perez N, Allamand V, Panthier JJ, Rousset JP (2004) *Gene Ther* 11:619–627.
34. Salas-Marco J, Fan-Minogue H, Kallmeyer AK, Klobutcher LA, Farabaugh PJ, Bedwell DM (2006) *Mol Cell Biol* 26:438–447.
35. Alkalaeva EZ, Pisarev AV, Frolova LY, Kisselev LL, Pestova TV (2006) *Cell* 125:1125–1236.
36. Chavatte L, Frolova L, Laugaa P, Kisselev L, Favre A (2003) *J Mol Biol* 331:745–758.
37. Ma B, Nussinov R (2004) *J Biol Chem* 279:53875–53885.
38. Vorobiev Yu, Kisselev L (2007) *Mol Biol (Moscow)* 41:103–111.
39. Moreira D, Kervestin S, Jean-Jean O, Philippe H (2002) *Mol Biol Evol* 19:189–200.
40. Sarkar G, Sommer SS (1990) *BioTechniques* 8:404–407.
41. Cassan M, Delaunay N, Vaquero C, Rousset JP (1994) *J Virol* 68:1501–1508.
42. Bidou L, Stahl G, Hatin I, Namy O, Rousset JP, Farabaugh PJ (2000) *RNA* 6:952–961.
43. Nguyen VT, Morange M, Bensaude O (1988) *Anal Biochem* 171:404–408.